REPORTING VICTIMISATION IN THE CRIME SURVEY FOR ENGLAND & WALES

A review by Joel Williams (TNS BMRB)
July 2016

TNS BMRB

Executive summary and conclusions

ONS uses the Crime Survey for England & Wales (CSEW) to report (i) the twelve-month *prevalence* of 'victimisation' within the relevant population (individuals aged 16+, individuals aged 10-15, or households) and (ii) the twelve-month *total number* of incidents of victimisation experienced by that population.

Reporting the *prevalence* of victimisation is relatively straightforward as survey respondents are coded either as a victim of the specified crime type in the last twelve months, or not. Reporting the *total number of incidents* of victimisation is more complex due to the existence of 'series' victimisations that comprise multiple incidents that are very similar, done under the same circumstances, and probably by the same people. Evidence from the survey methods literature suggests a substantial potential for measurement error when large incident counts are reported for a single victimisation.

Furthermore, because only a small subset of respondents report very high incident counts, the estimate of the total number of incidents for some crime types is *strongly* affected by random sampling error.

This combination of (potential) measurement error and substantial random sampling error led to a decision (in 1982, at the inception of the survey) to 'cap' incident counts to a maximum of 5 per victimisation for *all* crime types, regardless of distributional differences between them. Official statistics about the total number of incidents of victimisation have been based on capped data ever since. Walby et al[1] and others have suggested that this cap has obscured genuine trends in victimisation and should be discontinued. As part of its own review of CSEW methods, ONS commissioned TNS BMRB to carry out an independent review of this topic.

In summary, although the current practice of capping the number of incidents within a 'series' victimisation to a maximum of 5 is effective at reducing random sampling error – albeit not to a level that would make the point estimates particularly 'reliable' - there is a clear risk that it introduces *additional* error (downward bias) given its substantial impact on the point estimate for some crime types. In particular, the raw data for violence, threats and sexual offences includes many examples in which the number of incidents exceeds 5. For these crimes at least, a cap threshold of 5 is too suppressive.

So why not entirely discontinue capping the number of incidents per victimisation? After all, although the data is certainly affected by *random* measurement error (for example, respondents randomly under or over estimating the number of incidents in a 12 month reference period), there is no specific evidence for the kind of *systematic* measurement error which would lead to biased point estimates. However, if it was to make this change, ONS would effectively be accepting that the absence of evidence for systematic measurement error is strong evidence of its absence. That would

---

[1] See http://eprints.lancs.ac.uk/72272/4/Violence_Society_Research_briefing_1.pdf and http://www.civitas.org.uk/archive/pdf/CivitasReviewJun07.pdf

be a contentious position to take in any circumstance but in this case the use of uncapped data comes at a considerable cost. The *random* error (both sampling error and measurement error) affecting the estimates would be increased by an order of magnitude. ONS's ability to detect change in incident volumes from year to year would be severely curtailed, forcing it to report on multiple year aggregations of the data instead of single years as at present and, even then, reporting with reduced precision. Although one might argue that it is better to be imprecise but accurate in the long run, in practice, this would be a frustrating state of affairs. ONS would have to deal with huge random swings in incident volumes between reporting periods, obscuring the real change it seeks to detect and making communication of the results more difficult.

Bearing all this in mind, TNS BMRB suggests that a *pragmatic* approach is taken if an incident count is to remain part of the statistical set published by ONS.

- Firstly, multiple year aggregations ought to be used for reporting the number of victimisation incidents *if* the annual random sampling error is above a pre-specified threshold.
- Secondly, while it would be sensible to cap the raw data somewhat, TNS BMRB recommends a *lighter* cap, at the crime-specific 98th or 99th percentile point. As an illustration, this would mean capping violence at 36 series incidents per year rather than 5 if the 99th percentile was adopted as the cap threshold.
- TNS BMRB also recommends more prominent use of confidence intervals to display the true level of uncertainty surrounding the point estimates.

However, it is arguable whether the practice of reporting the number of victimisation incidents, even if made more reliable, is the *right* practice. An alternative would be to track the prevalence not just of victims of particular crimes but of individuals (or households) *suffering the same crime type more than once within the reference period of twelve months*. This would highlight the issue of repeat victimisation more directly than a simple sum of incidents which obscures the distribution of those incidents among the population. In addition, the point estimates for these statistics would be more reliable than any of the point estimates for incident totals discussed in this review.

In practice, the demand for estimates of the number of victimisation incidents is unlikely to go away, despite the problems of measurement. However, ONS could prioritise the prevalence estimates over the incident estimates when it reports on CSEW data. Given their greater reliability, this seems a sensible way forward.

Introduction: Reporting the prevalence and incidence of crime using the CSEW

ONS uses the Crime Survey for England & Wales (CSEW) to report (i) the twelve-month *prevalence* of 'victimisation' within the relevant population (individuals aged 16+, individuals aged 10-15, or households) and (ii) the twelve-month *total number* of incidents of victimisation experienced by that population.

Reporting the *prevalence* of victimisation is relatively straightforward as survey respondents are coded either as a victim of the specified crime type in the last twelve months, or not. Although there is a risk of measurement error here, the evidence from this and other crime surveys suggests the risk is low. Furthermore, the binary nature of this variable means that the random sampling error affecting the estimate of prevalence is tolerable for most crime types[2].

Reporting the *total number of incidents* of victimisation is more complex due to the existence of 'series' victimisations that comprise multiple incidents that are very similar, done under the same circumstances, and probably by the same people. Evidence from the survey methods literature suggests a substantial potential for measurement error, especially when large incident counts are reported for a single victimisation. Furthermore, because only a small subset of respondents report very high incident counts, the estimate of the total number of incidents for some crime types (particularly violence, threats and sexual offences) is *strongly* affected by random sampling error.

This combination of (potential) measurement error and substantial random sampling error led to a decision (in 1982, at the inception of the survey) to 'cap' incident counts at 5 per victimisation for *all* crime types, regardless of distributional differences between them. Official statistics about the total number of incidents of victimisation have been based on capped data ever since. Walby et al[3] and others have suggested that this cap has obscured genuine trends in victimisation and should be discontinued. As part of its own review of CSEW methods (including consultation with the GSS Methodological Advisory Committee) ONS commissioned TNS BMRB to carry out an independent review of this topic.

The TNS BMRB review examines alternatives to the current method of reporting the twelve-month total number of incidents of victimisation (for a given population). Most of these alternatives are alternative ways of capping the incident data that more fully respect the differences between crime types. However, TNS BMRB's favoured solution is to acknowledge the measurement and random sampling errors affecting this data and provide an alternative that better highlights repeat victimisation *and* has lower levels of measurement and random sampling error.

This could be achieved simply by extending the measure of prevalence, distinguishing those experiencing single incidents from those experiencing multiple incidents (either multiple incidents in the same victimisation or two or more 'separate' victimisations). This would be a simple way of

---

[2] The exception is for crime types that are rare within the population and consequently 'represented' by only a small number of victims within the CSEW sample.

[3] See http://eprints.lancs.ac.uk/72272/4/Violence_Society_Research_briefing_1.pdf

acknowledging that the experience of multiple incidents is qualitatively different from the experience of a single, probably isolated incident. Distinguishing single from multiple incidents ought to be straightforward with little additional risk of measurement error on top of that for the basic prevalence estimate. Furthermore, the random sampling error for percentages of this type should also be tolerable as it is effectively equivalent to capping at 2 the number of incidents within a victimisation but – crucially - without adding distortion to the population estimate, given that it would be a prevalence estimate rather than an estimate of a total.

Structure of this review

This review has several sections. In part 1, we describe the measurement process within the CSEW and evaluate the risk of measurement error. In part 2, we demonstrate how random sampling error affects the estimates and how this varies as a function both of sample size and of the population distribution of incidents of victimisation. In part 3, we describe a set of alternative methods of estimating the total number of victimisation incidents. In part 4, we evaluate these methods empirically in an effort to find a 'best fit' solution. In particular, we will focus on methods that provide sufficiently reliable estimates for us to detect genuine change over time. Part 5 draws these strands together and provides a set of recommendations.

Part 1: The measurement process within the CSEW and the risk of measurement error

For the CSEW, the respondent first completes an interviewer-mediated 'screener' module which gives him/her several opportunities to report experiences that might be classified as victimisations and that occurred within the last 12 months. Although each item within the module seeks information about a particular subset of crime types, the 'crime type' codes that arise from this module are preliminary only. The final codes are allocated based on the evidence from a detailed 'victim form' that the respondent completes for each (potential) victimisation. A team of experienced research staff assign these codes and there is a programme of 'blind' double-coding to ensure consistency in this respect.

Within the victim form, survey respondents are asked to place all (potential) victimisations within one of several time periods, mostly of three months' duration but including one 'ineligible' time period for placing victimisations that, on reflection, probably took place before the twelve month 'eligible' timeframe. This is intended to reduce the risk of respondents 'telescoping-in' incidents from outside the twelve month timeframe[4]. In combination, these methods probably ensure accurate coverage of the vast majority of victimisations occurring within the twelve month timeframe.

However, difficulties arise when a respondent reports that a victimisation comprised multiple incidents that were very similar, done under the same circumstances, and probably by the same people. These 'series' victimisations are represented by a single victim form, in which the detail is based on the *last* incident but where the respondent also estimates the number of incidents within the series (again, using the three month time periods described above). Each respondent provides

[4] The accuracy of the specific sub-frame allocation *within* the eligible period is likely to be lower than the accuracy of allocation within the twelve month time-frame as a whole.

an incident total between 1 and 96 for each three month time period with code 97 used for 97, 98+ or ‘too many to count’ (which may or may not be higher than 96). Consequently, the highest possible value over the full 12 month timeframe is 388 (97*4).

There is an obvious risk of measurement error at this point because it has been observed in several contexts that survey respondents tend to *approximate* if the real number is ten or more. Researchers in the US, working on the National Crime Victimization Survey, were able to demonstrate the implications of this by asking the same question at two different points within the victim form. Although the majority (84%) of those reporting a figure greater than ten reported the same figure the second time around, the minority reporting different figures often reported *very* different figures:

“…the magnitude of the discrepancy increased as the victims’ responses to the initial question increased. For example, the mean discrepancy between the two estimates for victims who initially reported 10 or more incidents was 5.5 victimizations. For those who initially reported 18 or more victimizations, the mean discrepancy was 8.4 victimizations. For those who initially reported 60 or more victimizations, the average discrepancy was 15.2 victimizations.”[5]

Measurement error is introduced by both (i) the natural need for the respondent to approximate when the true number of incidents is high, and (ii) the need to convert from the respondent’s natural way of estimating volume to the imposed way (a count of incidents). We may think we can spot some errors – e.g. a report of 96 bicycles stolen from a single household over the course of just three months seems obviously wrong– but other reports may seem implausible only because we have limited understanding of others’ lives.

Ultimately, we have no way of judging how much measurement error there is and – just as important - no way of judging how much of this measurement error is randomly distributed around the unobserved true number of incidents for each respondent in the sample, and how much of this error is *systematic*. Systematic error is more problematic and will occur when the measured number of incidents for the whole sample (or for a defined subsample) is significantly different from the unobserved true number. Systematic measurement error will tend to lead to systematic bias in point estimates that are derived from the data. Consequently, while random measurement error is often tolerated, systematic measurement error usually is not.

Our inability to quantify and decompose the measurement error is a critical problem for judging the methods used to produce official statistics on this topic. For example, one way of assessing alternative methods of estimating the total number of victimisation incidents would be to combine the measurement error (systematic and random) with the easy-to-calculate random sampling error to obtain an estimate of *total error*[6]. Without a reliable estimate of measurement error, there is no

---

[5] See Lauritsen et al (2012), *Methods for Counting High-Frequency Repeat Victimizations in the National Crime Victimization Survey*, Bureau of Justice Statistics, U.S. Department of Justice.

[6] Total error is equivalent to the expected (absolute) difference between the sample-based estimate and the true population value.

way to calculate this total error and therefore no way of objectively distinguishing between estimation methods.

However, although we cannot *definitively* select the best method for estimating the total number of victimisation incidents, it seems more reasonable to select a method which reflects the actual distribution of reported incidents than to select a method that places a uniform – and arbitrary - cap on those reports. Some additional random sampling error might be acceptable in exchange for additional sensitivity in this respect[7].

Part 2: Random sampling error

Random sampling error is a product of taking a sample from a larger population. Different samples drawn at the same time and in the same way will nevertheless produce different estimates because the actual set of people within each sample will differ in unobservable respects. The strongest influence on the *scale* of the random sampling error is sample size. Ignoring the complicating features of the sample design, the random sampling error for the total number of victimisation incidents (X) is equal to:

$SE(X) = NS/\sqrt{n}$

(SE = standard error = random sampling error; N = population size; S = standard deviation of total number of incidents per person (estimated by s, the standard deviation in the sample); n = sample size)

Random sampling error may be reduced by stratifying the sample frame first and drawing mini-samples from each stratum or by post-stratifying the responding sample so that any sample will have the same distribution with regard to gender, age, region and possibly other characteristics for which reliable population data is available. In practice, both strategies are employed for the CSEW and other surveys of similar design.

The impact of a random sampling error is not so much due to its absolute scale but due to its *relative* scale. If the random sampling error of a point estimate[8] is large relative to the point estimate itself, then its impact is large. Another random sampling error that is technically larger – in an absolute sense - but smaller relative to its point estimate will have a smaller impact.

Within this review, random sampling error is expressed as a 'coefficient of variation' (CV) – the standard error of the point estimate *expressed as a proportion of that estimate*. This CV value is a useful expression as it allows comparison across crime types with different (estimated) population distributions. It also gives an indication of our ability to detect real change in the number of crimes experienced by the population. For example, if the annual CV value is 5%, then a change of 20% in

---

[7] Having said that, if – for a particular crime type - the more sensitive method produces very similar population estimates to the 'insensitive' method of a uniform cap of 5, then it makes sense to retain the latter method because no revision of already published CSEW-based population estimates would be required.

[8] The term 'point estimate' includes totals (like this one) as well as means and proportions.

the number of incidents of this type will usually lead to us observing a significant difference in the survey data[9].

Although there will be some modest year-by-year variation in the number of crimes experienced by the population, our interest is in detecting the more substantial differences that may be evaluated as 'trends' over a longer time period. If the random sampling error is too great, these will not be detected even with the very large sample sizes typical of the CSEW (35,000 per year at present). If we were to assume that the number of crimes of a particular type will change only modestly (<10%) from year to year, then the ideal annual CV value would be no more than 2-3%.

Annualised estimates drawn from multiple years of data are likely to be much more reliable than single-year estimates because the multi-year random sampling error will be much smaller than the annual random sampling error - and more stable too - but, naturally, that does not improve our ability to detect *annual* change.

Technical note on estimating random sampling error

Estimating the random sampling error of each count for each estimation method may be done via the industry standard 'Taylor Series Expansion' (TSE) method or via more empirical alternatives (e.g. bootstrapping or Bayesian estimation) that are generally more appropriate for count distributions that do not follow the Poisson distribution. However, our experience is that bootstrap/Bayesian[10] credibility intervals tend to be symmetrical and of similar scale to the TSE confidence intervals *whatever the shape of the variable's distribution within the sample* so long as the statistic is an unconditional count, mean or proportion and the sample size large. We have used standard TSE estimates for this study since these are used more generally for CSEW-based point estimates[11].

These TSE estimates account for the stratification and clustering of the sample as well as the weighting of the data. For simplicity, the various changes of design over the course of the data period 2003-15 are ignored in this review. In particular, police force area has been used as the stratum identifier for the whole period rather than any lower level stratum specific to particular sub-periods. We have also not applied the finite population correction that should be applied to the between-cluster component of the random sampling error for some strata over the 2008-12 data period and for all strata over the 2012-15 data period. These two simplifications mean that the random sampling error estimates quoted in this report will be slightly larger than their true values (by an average of around 10% for the 2012-15 data period).

[9] More formally, this means that we have an 80% probability of recording a 'significant' difference (if the critical p value is set at the standard of .05) if there was a real life 20% change in the number of victimisation incidents of that type.

[10] This holds if an uninformative prior probability distribution is selected *or* an informative one is selected but one with relatively wide error bounds.

[11] The term 'point estimate' includes sums, means and proportions.

Part 3: Alternative methods for estimating the total number of victimisation incidents

Official statistics have a number of dimensions of quality[12], one of which is that the method for estimating a quantity should be *transparent* and reliant on a minimum number of broadly accepted assumptions. In the survey context, the requirement of transparency has led to a preference for design-based estimation over model-based estimation. In fact, for surveys like CSEW, ONS *does* employ models of response propensity as part of the estimation system but its broad objective is to publish estimates that are independent of contestable modelling assumptions and that are comprehensible to the intelligent layperson.

Consequently, any change in the estimation process must not make the process of statistical estimation more obscure than it is now. In our view, this reduces the options for estimating the number of victimisation incidents to the following:

a) Annual count with a uniform respondent cap threshold of ***y*** same-perpetrator incidents for any single crime type (i.e. a generalisation of the current method in which y=5)

b) As a) but with different cap thresholds for different crime types

c) An annualised average of the *un*capped counts for each crime type, as recorded in the ***t*** most recent years of data

d) As d) but with different values of ***t*** for different crime types

e) Instead of counts, track the proportion of the population with different *ranges* of counts (e.g. 0; 1-***y***; >***y***)

There are other methods that might be envisaged but the practical consequence of adopting one of them would be very similar to one or other of the methods listed above. For example, we considered an alternative to capping counts at a threshold level ***y***. Instead of substituting reported counts above ***y*** with the value ***y***, we could substitute them with the unweighted median count for the subsample reporting ≥***y*** incidents. In theory, this method would combine the robustness of a single threshold level based on historical data with the flexibility to respond to temporal changes in the reporting distribution. However, in practice, this is not so very different from selecting methods (a)or (b) but with a higher value for ***y***.

With the caveat that the list of methods under consideration is not exhaustive, we now discuss each in more detail.

Method a) is a generalisation of the method currently used to report CSEW-based incident totals, in which ***y***=5 is used as the cap threshold. Retaining this method (with y=5) has the considerable advantage of coherence with all published CSEW-based population estimates so no revision of these would be required if this method is retained. However, the justification for selecting ***y***=5 in 1982 is

[12] See the UK Statistics Authority (2009) *Code of Practice for Official Statistics, v1.0* , accessible at https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-codeofpracticeforofficialstatisticsjanuary2009_tcm97-25306.pdf

unknown and the empirical conditions that held in 1982 may no longer hold in 2016. Furthermore, the assumption that a uniform value for ***y*** – of whatever level - is appropriate for *all* types of crime is hard to justify given the systematic differences found in the raw data.

Method b) differs from method a) by utilising a different value of ***y*** for each crime type, thereby accommodating the different incident distributions observed for different crime types rather than imposing a constraint of uniformity as with method a). For this project, we have experimented with a value of ***y*** equal to the crime-specific 98th or 99th percentile values among all reported series incidents over the entire data period (April 2003 to March 2015)[13]. For most crime types, this value for ***y*** is sufficiently different from 5 – and from the 'uncapped' maximum – to generate annual incident totals that are distinctive from those produced by either of these other methods.

Method c) was recommended by Walby et al as a way of reducing the random sampling error of unadjusted estimates but it sacrifices some of the timeliness, adding a lag equal to (***t***-1)/2 years. Her work was focused on violent crimes and she suggested that ***t***=3 would be suitable. Method c) may be refined into method d) by setting different (integer) values of ***t*** as appropriate for each crime type. This adds flexibility – and is defensible statistically – but presents a greater communication challenge, and is less transparent to users, than the more uniform approach.

Method e) is suggested as an alternative to the paradigm of reporting an annual number of victimisation incidents. Instead, ONS may track the proportion of the population experiencing *multiple* incidents of a particular crime over the preceding twelve months as well as tracking the proportion of the population experiencing *any* incidents. Arguably, this would bring a sharper focus on the distribution of incidents within the population than is possible within the count paradigm.

Part 4: Evaluation of alternative methods of estimation

The results from this evaluation are arranged thematically. In section 1, we assess the distribution of victim form incident counts over the whole data period (2003-15). This is useful for demonstrating the suitability or otherwise of a uniform cap threshold on the number of incidents per victim form as well as showing where the crime-specific 98th or 99th percentile cap thresholds would fall (method (b)).

In section 2, we assess the random sampling error we could expect under each of methods (a), (b), (c) and (d).

In section 3, we assess the expected difference in the *level* of the estimates under each method. Because we cannot definitively say which method yields the highest net measurement accuracy, this is included largely to understand the impact of changing method on the victimisation incident totals reported by ONS.

[13] Using percentile values as cap thresholds has an empirical advantage over arbitrarily selected thresholds. However, so long as the selected cap thresholds respect the distribution of the data, alternative values might be proposed instead. The 98th and 99th percentiles should be treated as *illustrative* only.

In section 4, we briefly assess alternatives to reporting the number of victimisation incidents (including method (e)) and in section 5 we draw together our conclusions and make a preliminary recommendation for ONS.

Data period and selected estimates

For this evaluation, we have used all CSEW data from April 2003 to March 2015 inclusive. The number-of-incidents variables under consideration are aggregations from single crime types that are used in ONS reporting[14], namely:

- *Individual-level crimes (if aged 16+)*
    - Violence excluding robbery and sexual offences (12 year number of victim forms = 13,035)
    - Robbery (12yr n = 1,951)
    - Personal theft (12yr n = 5,327)
    - 'Other' personal theft (12yr n = 9,569)
    - Sexual offences (12yr n = 599)
    - Threats (12yr n = 12,056)
- *Household-level crimes*
    - Burglary (12yr n = 17,408)
    - 'Other' household theft (12yr n = 21,002)
    - Theft of/from motor vehicle (12yr n = 25,329)
    - Bicycle theft (12yr n = 8,357)
    - Vandalism (12yr n = 35,977)

Results

1. *The distribution of victim form incident counts over the period 2003-15*

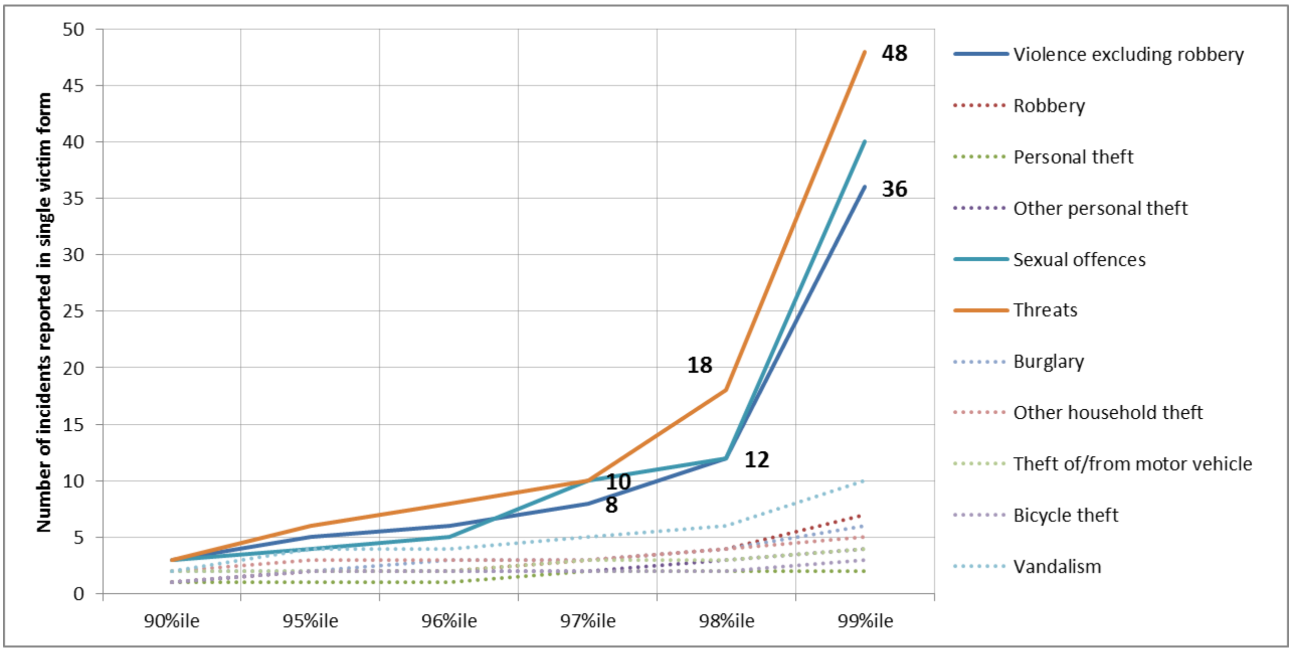
Chart 1 shows the (weighted) percentile values for the number of incidents per victim form for each crime type in this study. The numbers for violence, sexual offences and threats are picked out to highlight their distinction from the other crime types.

[14] See Appendix 2 of the CSEW User Guide for a list of single crime types for each aggregation: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/crime-statistics-methodology/user-guides/user-guide-to-crime-statistics.pdf

The chart demonstrates the very considerable variation between crime types in the distribution of incident totals per victim form. For some crime types – represented with dotted lines – capping the incident total at a maximum of 5 per victim form has minimal effect on any derived statistics. For personal theft, 'other' personal theft, 'other' household theft, theft of/from a motor vehicle and bicycle theft, a cap threshold of 5 is at or above the 99th percentile.

For others, the cap threshold of 5 has a stronger impact. A cap threshold of 5 is only around the 95th percentile for violence (excluding robbery) and for threats, and around the 96th percentile for sexual offences. A cursory glance at the 99th percentile value for these three crime types shows counts ranging from 36 to 48. Even the 98th percentile value ranges from 12 to 18 for these three crime types. Enforcing a cap threshold of a maximum of 5 incidents per victim form represents a significant suppression of the raw data. On the face of it, using a uniform cap (method (a)), especially of just 5, does not seem justified for these crime types. The primary focus for the remainder of the review will be on the three crime types that are most strongly affected by this cap threshold: violence, threats and sexual offences.

Chart 1: The distribution of victim form incident counts over 2003-15



2. An assessment of random sampling error affecting annual crime incident count estimates

Chart 2 shows the annual random sampling error we might expect from adopting method (b) - at the 98th and 99th percentile points – in place of the uniform cap threshold of 5 and also in comparison with an 'uncapped' estimation method.

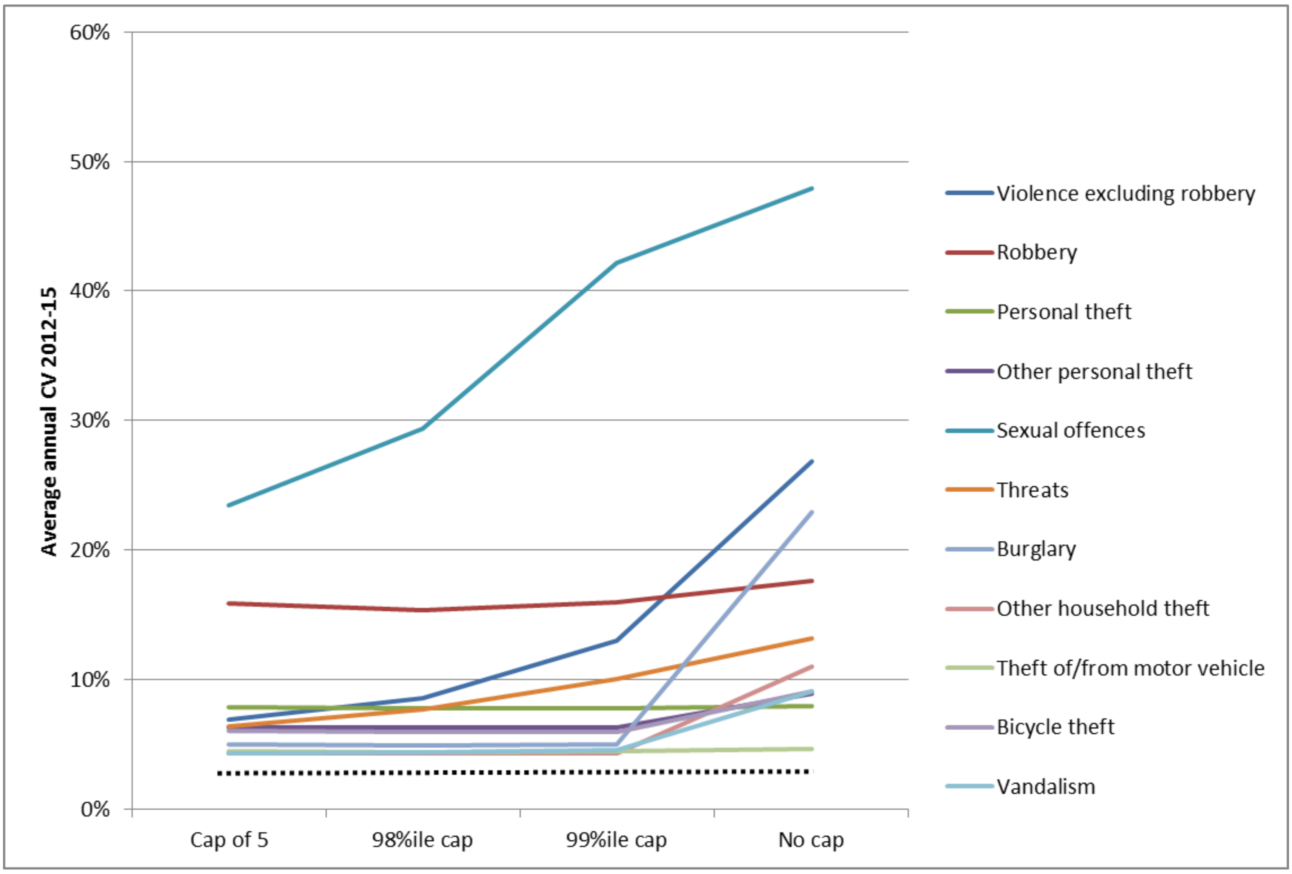
As noted in Part 2, within this review we have expressed random sampling error as a 'coefficient of variation' (CV) – the standard error of the total number of victimisation incidents *expressed as a proportion of that estimate*. This CV value is a useful expression as it allows comparison across crime types with different (estimated) population distributions. As also previously noted, if we hold a prior assumption that the number of victimisation incidents of a particular type will change only modestly (<10%) from year to year, then the ideal annual CV value would be no more than 2-3% as that would allow ONS to formally 'detect' changes of that magnitude.

Chart 2 is based on data from 2012-15 (when the overall household sample size was fixed at 35,000 per year), averaging the annual CVs for each of the three component twelve-month periods (2012-13, 2013-14, and 2014-15). It shows that – even with a uniform cap of 5 incidents per series victimisation – annual CV values are higher than the 2-3% ideal (shown with a black dotted line) for *every* crime type. In some cases, the annual CV value is several times greater than this ideal.

With the exception of robbery, personal theft and theft of/from a motor vehicle, the random sampling error is *much* higher with the uncapped estimation method than it is with the current estimation method based on a uniform cap of 5 incidents per victim form. The random sampling error ratio between the two estimation methods goes as high as four or five to one (burglary and violence) although a ratio of two to one is more typical. To put this in particularly stark terms, this level of random sampling error would be expected from the uniform-cap estimation method if the sample size was *one quarter* of what it actually is. To use uncapped incident reports in official estimation, ONS would have to be very confident that the high value counts are not affected by systematic measurement error and that the inflation in random sampling error is a worthwhile exchange.

For most crime types (robbery, personal theft, 'other' personal theft, burglary, 'other' household theft, theft of/from a motor vehicle and bicycle theft), the difference in random sampling error between the current estimation method (with a uniform cap of 5 incidents per victim form) and an alternative 98th or 99th percentile cap is negligible because the 99th percentile is not (much) higher than 5. For other crime types, the difference is much larger, particularly for violence, sexual offences and threats for which a 99th percentile cap would mean a near doubling of the random sampling error compared to a cap of 5. In all three of these cases, the 98th percentile cap yields a random sampling error somewhere between the two but closer to that of the cap of 5 than to that of the 99th percentile cap (which is a much higher value for these three crime types).

Chart 2 Relative random sampling error (coefficient of variation) for annual numbers of victimisation incidents (2012-15) of each crime type under four estimation methods



An alternative to the practice of publishing the number of victimisation incidents based solely on that year's data is method (c) – and its extension method (d) - as suggested by Walby et al. With this method, ONS would effectively annualise[15] the last ***t*** years of uncapped data, sacrificing some timeliness in exchange for a reduction in random sampling error which, in turn, would allow a less suppressive use of the raw victim form data. All things being equal, two years of data should have 71% of the random sampling error affecting one year of data, and three years of data should have 58% of a single year's random sampling error[16]. However, even with three years of data, the random sampling error affecting this 'annualised' estimate would – for several crime types - still be slightly higher than the random sampling error affecting the equivalent cap-5 *single* year estimate. Consequently, the problem of detecting change over time would remain. Indeed, it would be exacerbated due to the greater 'lag' in the estimate that follows from aggregating together multiple years of data.

---

[15] An annualised estimate is just the average over *t* years of data but scaled to fit with the population totals used for the last year in the set of *t* years. This annualising approach is only suitable from the inception (in 2000) of annual – and then continuous – data collection.

[16] Two years of data will have half the sampling variance of one year of data, and the standard error (called 'random sampling error' in this review) will be *the square root* of half the size, i.e. will be √1/2 = 71% as large. For three years of data, the standard error will be √1/3 = 58% as large.

There is another problem with this approach too. The reduction in random sampling error is not a fixed 29% for a two year aggregation and 42% for a three year aggregation. It varies by time period and by crime type. Single year CV values are rather volatile and that volatility carries through to some extent to the multi-year aggregations. A good example is violence which has single year CV values of 22% (2012-13), 40% (2013-14) and 18% (2014-15), averaging at 27% over that period. The very large value of 40% for 2013-14[17] contributes to CV values of 17% (2011-14) and 16% (2012-15) for the three year aggregations that contain 2013-14. These CV values are only slightly lower than the 'typical' single year value.

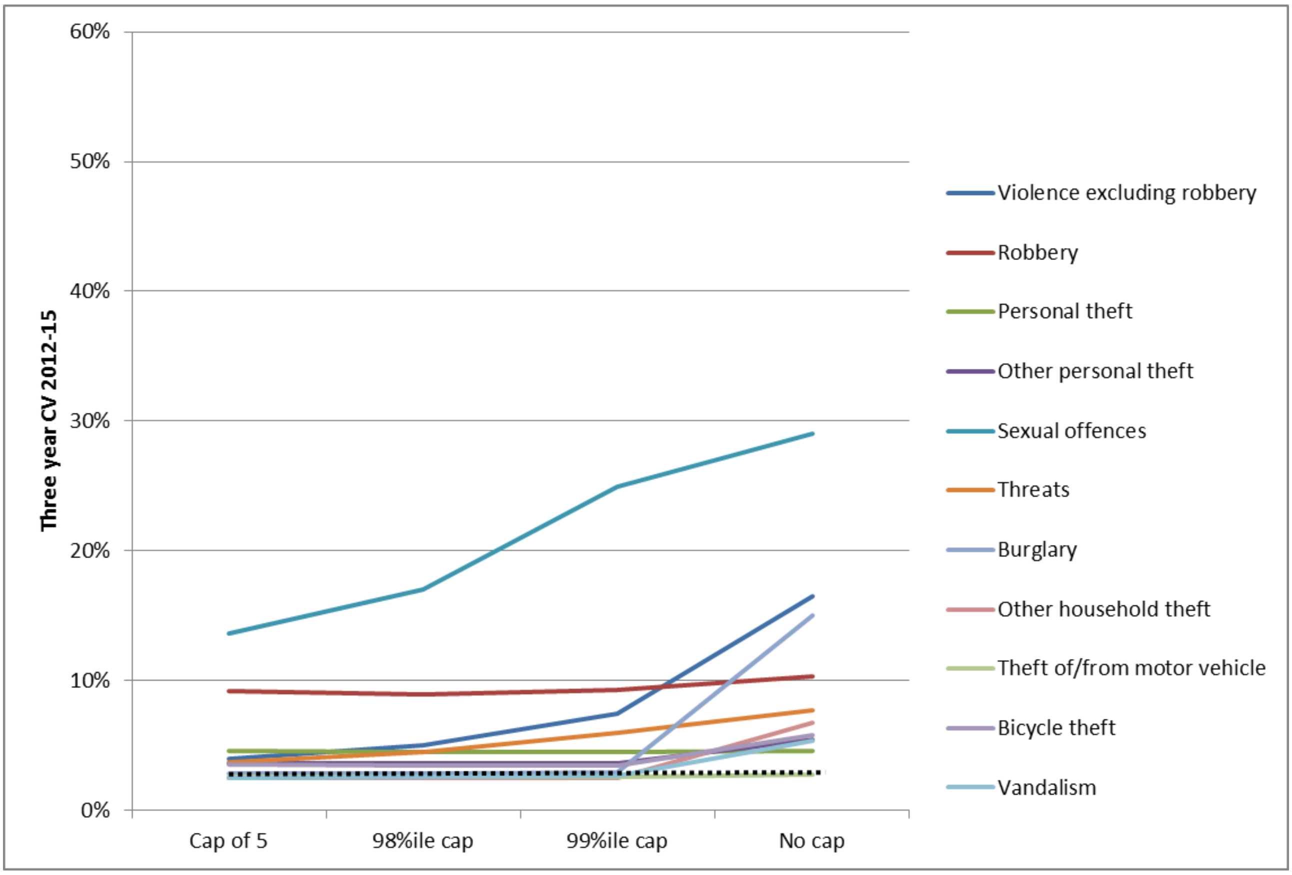
The multi-year random sampling error for other crime types is much more aligned with expectations, largely because there tend to be few victim forms with high incident totals. However, even here there are exceptions. A small number of (unrealistic?) reports of bicycle theft in 2014-15 strongly influence the level of random sampling error in aggregates that contain this year. The CV value jumps from 3.3% in 2011-14 to 5.8% in 2012-15 and this will continue for two more aggregations until this anomalous year drops out of the reporting dataset.

Given these drawbacks, it would make sense to develop this approach so that it is based not on multi-year aggregation of *uncapped* victim form data but on multi-year aggregation of *lightly-capped* victim form data (e.g. at the crime-specific 98th or 99th percentile values). The gain with respect to random sampling error would be much more predictable and – in general – the random sampling error would also be much lower.

For illustration, Chart 3 shows the level of random sampling error under each candidate estimation method for 2012-15 if ***t***=3.

[17] This was due to a higher than usual number of very large 'series' incident counts.

Chart 3 Relative random sampling error (coefficient of variation) for 3-year annualised numbers of victimisation incidents (2012-15) of each crime type under four estimation methods



An approach based around the 99th percentile cap threshold would lead to CV values that are – with the exception of sexual offences - low enough to regard the estimate of the number of victimisation incidents as tolerably precise. The same is also true of an approach based around the 98th percentile but, for illustration purposes, we focus on the 99th percentile because it more strongly respects the underlying shape of the data.

Method (d) suggests using a value for ***t*** that is sensitive to the different random sampling error levels observed for each crime type. One way of incorporating this in a systematic way that is comprehensible for users would be to select ***t*** such that the expected CV value is below a fixed threshold but with the constraint that ***t*** cannot exceed 3 (to avoid too much lag in the estimate). This CV threshold might be set at c5% so that the 95% confidence interval does not exceed +/-10% of the point estimate and also to ensure a reasonable probability of detecting substantial change in the number of victimisation incidents from one period of ***t*** years to the next (potentially overlapping) period of ***t*** years[18].

Table 1 shows the CV values under this approach for each crime type. The suggested value for ***t*** for each crime type is given in column 2. For four crime types (violence, robbery, sexual offences and

[18] The ideal of a CV value of 2-3% does not seem achievable while respecting the crime-specific distributions in the number of incidents per victim form.

threats) the three-year CV value – at least for the 2012-15 period - exceeds 5% so the published estimates would still lack sufficient precision to detect realistic levels of change over time. However, the CV is no greater than 5% for any of the other crime types and – with the exception of sexual offences – none is greater than 10%. It is reasonable to describe the point estimates under this estimation system as (i) tolerably respectful of the raw victim form data, and (ii) tolerably precise for publication.

Table 1 99%ile Capped CV values for optimal multi-year aggregation, 2012-15

	Optimum *t*	CV expectation (based on 2012-15 data)	2012-15 number of victim forms over *t* years (averaged where *t*<3)
Violence excluding robbery	3	8%	2,043
Robbery	3	9%	246
Personal theft	3	5%	1,107
Other personal theft	2	4%	1,232
Sexual offences	3	25%	135
Threats	3	6%	2,262
Burglary	1	5%	983
Other household theft	1	4%	1,384
Theft of/from motor vehicle	1	5%	1,186
Bicycle theft	2	4%	1,018
Vandalism	1	5%	1,711

Before we move on from the subject of random sampling error, it is worth considering an alternative to the explicit publication of a point estimate. ONS could instead publish a *plausible range of values* - perhaps based on 95% confidence intervals – rather than a specific number. This would better convey the uncertainty in the estimate - something that is not currently well understood by lay consumers of the data - and should also encourage caution among those reporting the figures. Providing the point estimate alongside the confidence interval may not achieve the same result, potentially relegating the confidence interval to a statistical caveat rather than the main estimate.

However, there would be nothing to stop reports of 'up to [95% confidence interval upper bound] crimes of type [x]' that ignore the lower bound. There is a risk that the spurious precision of a single incident total based on the sample data is replaced with a probably less accurate version of the same.

3. An assessment of the expected difference in point estimates by estimation method

As noted in several places, we cannot definitively point to an estimation method and declare it to have the least measurement error. Consequently, we are forced to make a more subjective assessment about which is the 'best' estimation method. It seems reasonable to prefer an estimation method which better reflects the differences between crime types with respect to incident distribution over the current uniform cap-5 estimation method. It is also reasonable to expect that point estimates have a maximum random sampling error before being quoted and that multiple-year aggregations are sensible for some crime types (particularly violence, robbery, personal theft, threats and sexual offences) in order to meet this goal.

Inevitably, switching estimation method would lead to changes from already-published point estimates so it is worth elaborating on the scale of change this would bring about. To keep this manageable, table 2 shows the annual number of victimisation incidents for 2014-15 under three estimation methods: the current single year uniform cap-5 method, an annualised 3 year 'uncapped' estimation method (as promoted by Walby et al) and this author's preferred estimation method in which a crime-specific ***t*** years of data is used and the victim form data capped at the crime-specific 99th percentile.

For most crime types the adjustment in the point estimate would be small and within the confidence interval of the published (cap-5) point estimate. However, there would be a substantial increase in the point estimates for violence, robbery, sexual offences and threats, all of which would increase by more than 50% under the method proposed above as a reasonable approach. Switching method would not only mean significant work to re-estimate past point estimates but would also necessitate a cogent explanation of why the new method is being adopted.

Table 2: Point estimates (total crime incidents in population 2014-15) by three different estimation methods

	(i) Single year; uniform cap of 5	(ii) Average for *t* years capped at 99%ile value	(iii) Uncapped 3yr average	(ii)-(i)	(iii)-(i)
Violence excluding robbery	1,321,247	2,267,796 (t=3)	3,578,499	+72%	+171%
Robbery	89,717	155,004 (t=3)	168,497	+73%	+88%
Personal theft	450,767	513,367 (t=3)	521,500	+14%	+16%
Other personal theft	741,157	880,322 (t=2)	937,841	+19%	+27%
Sexual offences	115,351	174,348 (t=3)	268,857	+51%	+133%
Threats	1,450,755	2,309,013 (t=3)	2,697,856	+59%	+86%
Burglary	785,428	825,716 (t=1)	1,184,030	+5%	+51%
Other household theft	986,424	1,079,496 (t=1)	1,274,346	+9%	+29%
Theft of/from motor vehicle	922,782	954,302 (t=1)	970,803	+3%	+5%
Bicycle theft	381,457	401,276 (t=2)	423,160	+5%	+11%
Vandalism	1,333,802	1,578,774 (t=1)	1,822,761	+18%	+37%

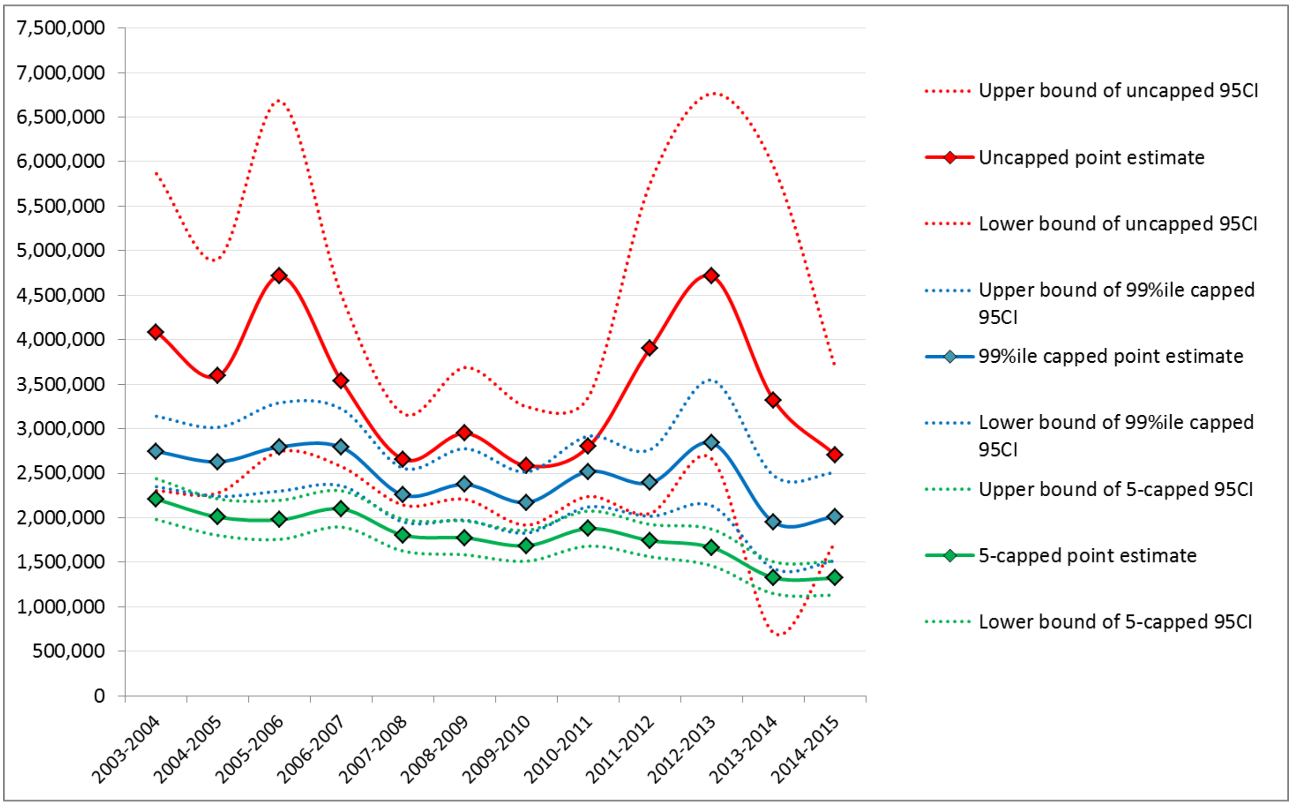
To further illustrate these findings, Chart 4 shows the *single year* estimate of the number of 'violence' victimisation incidents over the entire course of the 2003-15 data period. It includes the point estimate (solid lines) under (i) an 'uncapped' estimation method, (ii) a 99th percentile cap method (with the threshold set at 36) and (iii) the current method in which victim form data is capped at 5. Although the preceding text suggests that annualised multi-year data should be used for estimation purposes, we start the illustration with single year data (chart 4) before moving on to multi-year data (chart 5) to demonstrate the implications of retaining the current single year approach to estimation.

It should be immediately obvious from chart 4 that the estimation method has a substantial effect on not only the point estimate and the width of the 95% confidence interval but on the interpretation of change over time. The uncapped point estimate shows a frankly untenable degree of movement over this period. Overall, the average annual change is -1% but the average *absolute* annual change is +/-21%. Capping the victim form data at 36 incidents produces a more reasonable degree of movement but we still observe a drop of c890,000 incidents (-31%) between 2012-13 and 2013-14 which seems unlikely. The average absolute annual change is +/-11%, half that of the uncapped estimation method but still high given we earlier considered that a 10% change from one year to the next to be at the outer edge of expectation.

The standard 5-cap estimation method flattens out much of the variation from year to year with a fall of 340,000 incidents (-20%) between 2012-13 and 2014-15 the largest single year change in the point estimate, and an average absolute annual change of +/-7%. The 5-cap variant seems the most realistic method in terms of the degree of annual change, although it misses entirely the upwards movement between 2011-12 and 2012-13 recorded by the other two estimation methods (and may therefore have missed a genuine movement, even if later categorised as a 'blip').

However, although the 5-cap method might have the scale of change right, there is no ignoring the enormous difference in the point estimates produced by the three estimation methods with the 5-cap official estimate tending to be below the lower bound of the confidence interval of the other two methods. In other words, although the 5-cap method produces a reasonable degree of annual variation, the *level* is probably wrong unless there is a very substantial upwards-bias in the data from victim forms with more than 5 incidents.

Chart 4 Single year point estimates (number of violence victimisation incidents) under three estimation methods

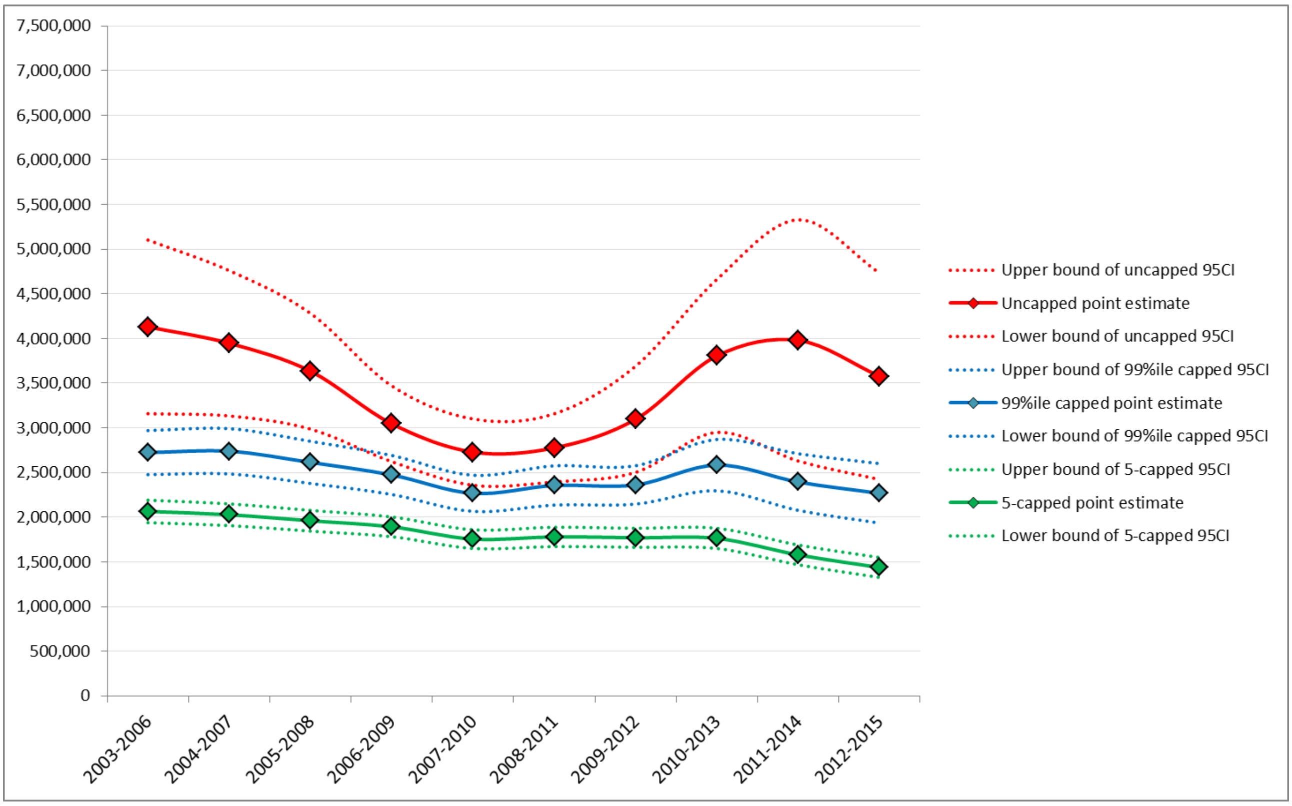


These effects are reduced if *annualised* three-year estimates are used instead of single year estimates. As Chart 5 shows, the between-year fluctuation in the point estimates is smoothed out; the short-lived increases in the single-year number of victimisation incidents (2005-6 and 2011-13) are blended in with other years to produce a longer-lasting but more modest effect. The average annual change with uncapped data becomes +/-10% (instead of +/-21% as with the single year estimation method), +/-5% with 99th percentile capped data (instead of +/-11%) and +/-4% with the current cap of 5 (instead of +/-7%).

The point estimates are also proportionately closer to each other. On average, the 99%ile capped estimate is 82% of the uncapped estimate (compared with 73% for single-year data) and the cap-5 estimate is 62% of the uncapped estimate (compared with 53% for single-year data).

These timelines look more 'sensible' than the single-year timelines: they fluctuate in a more realistic fashion but without changing the overall level of victimisation typical of each estimation method. However, as noted above, annualising three years of data does not solve the problem that each estimation method produces distinctly different levels of victimisation and that choosing between them is a matter of belief about the relative size of the two types of survey error: (i) the level of bias introduced by capping the data and (ii) the additional random sampling error that follows from using uncapped nor lightly-capped data.

Chart 5 Annualised (three-year) point estimates (number of violence victimisation incidents) under three estimation methods



4. Alternative approaches to the provision of a single count of crimes

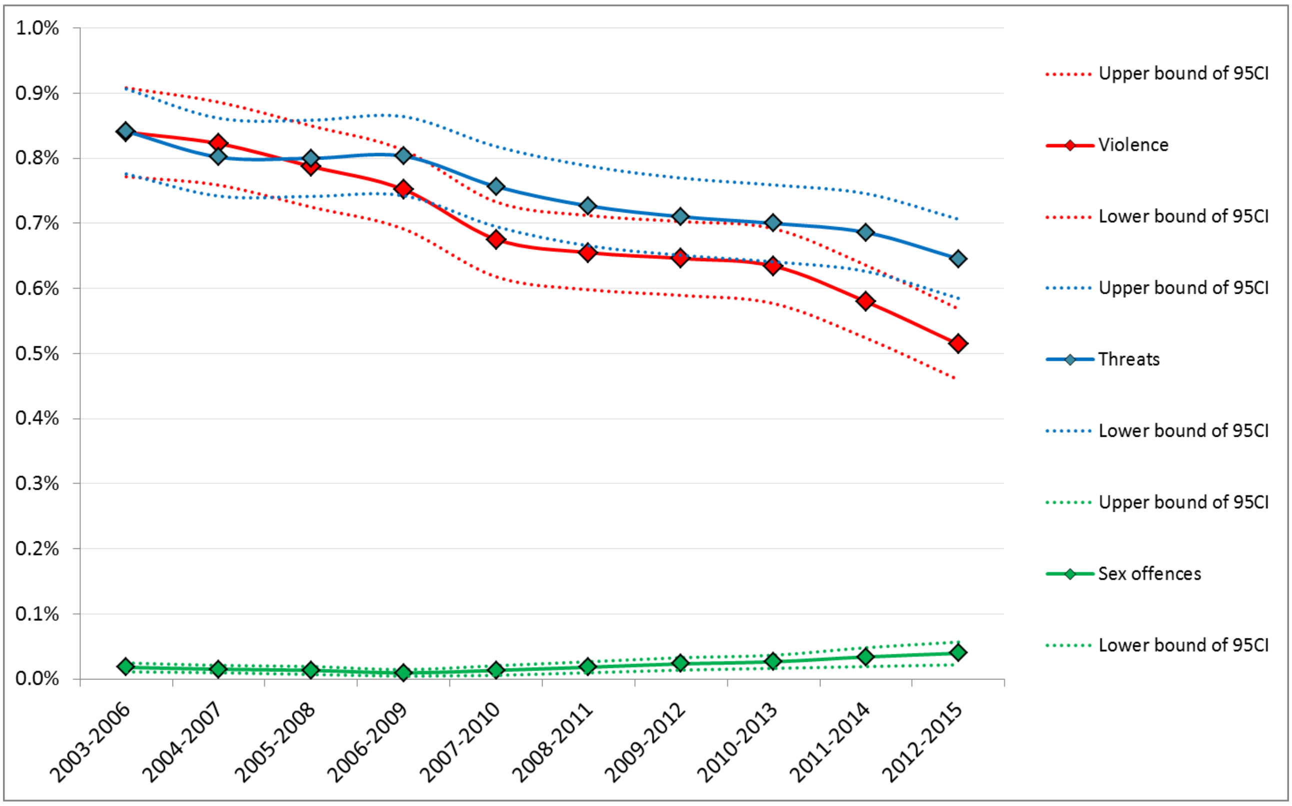
Given the quandary described above, consideration should be given to taking a completely different tack and adopting method (e). With method (e), we add to the reports of crime type prevalence by distinguishing those experiencing single incidents from those experiencing multiple incidents (either multiple incidents in the same series or two or more 'separate' (non-series) incidents) when reporting the CSEW results. This would be a simple way of acknowledging that the experience of multiple incidents is quite different from the experience of a single, probably isolated incident. The random sampling error for percentages of this type is smaller than for incident sums as it is effectively equivalent to capping incident totals at 2 per victim form but *without* adding distortion to the population estimate, given it would be a prevalence estimate rather than an incidence estimate.

Chart 6 shows the (annualised three-year) prevalence of repeat victimisation for three crime types - violence, threats and sexual offences – over the full course of the data period. For both violence and threats, these estimates of the prevalence of repeat victimisation would have CV values of around 4-5% (significantly less than for the current single year cap-5 estimate of the number of victimisation incidents for which the CV value is about 6-7%). Even *single*-year estimates of the prevalence of repeat victimisation would have CV values of only 8-10%. The CV value for repeat sex offence victimisation would still be very high, even with a running three-year estimate (over 20%). Given the

TNS

very small number of victim forms each year (<50, so <150 over three years), it does not seem sensible to imply much precision for *any* CSEW-based reports on sex offences[19].

Chart 6: Prevalence of repeat victimisation (2+ incidents in 12 month reference period) for three crime types (violence, sexual offences, threats): three-year annualised data



If this new approach was to be taken, the benefits of a switch of emphasis would need to be communicated carefully. The principal benefits are (i) that we can spotlight multiple and serial victimisation for the first time - a phenomenon that is not obvious when reporting the simple number of victimisation incidents – and (ii) that we drop – or at least deemphasise – the least reliable data (incident counts) reported as official statistics. One approach might be to use the method (e) estimate as the headline but also include, as a *secondary* estimate, annualised three-year incident totals using the confidence interval method described in section 3.

<u>5. Summary and conclusions</u>

In summary, this analysis suggests that reporting the number of victimisation incidents of a particular type in a particular year is fraught with difficulty. The point estimate is subject to strong random sampling error and (potentially) random and systematic measurement error too.

[19] This is especially true given the high likelihood of 'missing' victim forms due to non-disclosure. Although we can be confident that fewer than one in a thousand adults suffer repeat sexual offences within a twelve month time-frame, that is about *all* we can say.

The current practice of capping the number of incidents within a 'series' victimisation at 5 is effective at reducing random sampling error – albeit not to a level that would make the point estimates 'reliable' - but there is a risk that it introduces *additional* error given its substantial impact on the point estimate for some crime types. In particular, the raw data for violence, threats and sexual offences includes many examples in which the number of incidents exceeds 5. For these crimes at least, a cap threshold of 5 is too suppressive.

So why not entirely discontinue capping the number of incidents per victimisation? After all, although the data is certainly affected by *random* measurement error, there is no specific evidence for the kind of *systematic* measurement error which would lead to biased point estimates. However, if it was to make this change, ONS would effectively be accepting that the absence of evidence for systematic measurement error is strong evidence of its absence. That would be a contentious position to take in any circumstance but in this case the use of uncapped data comes at a considerable cost. The *random* error (both sampling error and measurement error) affecting the estimates would be increased by an order of magnitude. ONS's ability to detect change in incident volumes from year to year would be severely curtailed, forcing it to report on multiple year aggregations of the data instead of single years as at present and, even then, reporting with reduced precision. Although one might argue that it is better to be imprecise but accurate in the long run, in practice, this would be a frustrating state of affairs. ONS would have to deal with huge random swings in incident volumes between reporting periods, obscuring the real change it seeks to detect and making communication of the results more difficult.

Bearing all this in mind, TNS BMRB suggests that a *pragmatic* approach is taken if an incident count is to remain part of the statistical set published by ONS.

- Firstly, multiple year aggregations ought to be used for reporting the number of victimisation incidents *if* the annual random sampling error is above a pre-specified threshold.

- Secondly, while it would be sensible to cap the raw data somewhat, TNS BMRB recommends a *lighter* cap, at the crime-specific 98th or 99th percentile point. As an illustration, this would mean capping violence at 36 series incidents per year rather than 5 if the 99th percentile was adopted as the cap threshold.

TNS BMRB also recommends more prominent use of confidence intervals to display the true level of uncertainty surrounding the point estimates. This might include using the confidence intervals alone to discourage fixation on a spuriously precise point estimate. However, it is arguable whether the practice of reporting the number of victimisation incidents, even if made more reliable, is the *right* practice. An alternative would be to track the prevalence not just of victims of particular crimes but of individuals (or households) suffering the same crime type more than once within the reference period of twelve months. This would highlight the issue of repeat victimisation more directly than a simple sum of incidents which obscures the distribution of those incidents among the population. In addition, the point estimates for these statistics would be more reliable than any of the point estimates for incident totals discussed in this review.

In practice, the demand for estimates of the number of victimisation incidents is unlikely to go away, despite the problems of measurement. However, ONS could prioritise the prevalence estimates over the incident estimates when it reports on CSEW data. Given their greater reliability, this seems a sensible way forward.

Joel Williams (Head of Methods, TNS BMRB), July 2016